Response alternatives: the impact of their choice and presentation order

Schwarz, Norbert; Hippler, Hans-Jürgen

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Response Alternatives:
The Impact of their Choice and Presentation Order

Norbert Schwarz, Hans-J. Hippler

ZUMA-Arbeitsbericht Nr. 90/08

Zentrum für Umfragen, Methoden und
Analysen e.V. (ZUMA)
Postfach 12 21 55
D-6800 Mannheim 1

Seit Juli 1983 sind die ZUMA-Arbeitsberichte in zwei Reihen aufgeteilt:

Die **ZUMA-Arbeitsberichte (neue Folge)** haben eine hausinterne Begutachtung durchlaufen und werden vom Geschäftsführenden Direktor zusammen mit den übrigen Wissenschaftlichen Leitern herausgegeben. Die Berichte dieser Reihe sind zur allgemeinen Weitergabe nach außen bestimmt.

Die **ZUMA-Technischen Berichte** dienen zur hausinternen Kommunikation bzw. zur Unterrichtung externer Kooperationspartner. Sie sind nicht zur allgemeinen Weitergabe bestimmt.

The attached reprint replaces ZUMA-Arbeitsbericht No. 90/08 by the same authors.

N. Schwarz & H. J. Hippler (1991). Response alternatives: The impact of their choice and presentation order. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), Measurement error in surveys (pp. 41-56). Chichester: Wiley.

CHAPTER 3

RESPONSE ALTERNATIVES: THE IMPACT OF THEIR CHOICE AND PRESENTATION ORDER

Norbert Schwarz and Hans-Jürgen Hippler
Zentrum für Umfragen, Methoden, und Analysen (ZUMA)

3.1 INTRODUCTION

That the choice and presentation order of response alternatives in a survey can greatly influence the obtained results is no news and has been widely documented in the survey literature (cf. Payne, 1951; Sudman and Bradburn, 1974; Molenaar, 1982; Hippler and Schwarz, 1987). However, the underlying cognitive and communicative processes are not well understood, which makes it difficult to predict which effects may be expected under various conditions. In the present chapter, we will review a cognitive research program that explores the psychological processes that mediate the impact of response alternatives on respondents' reports. In addition to summarizing parts of our own research, we provide a selective review and conceptual integration of the available literature on the effects of choice and presentation order of response alternatives.

We begin with a comparison of open and closed question formats, focusing on the information that respondents extract from different types of response alternatives provided to them. Subsequently we explore the

The reported research was supported by grants Schw278/2 and Str264/2 from the Deutsche Forschungsgemeinschaft to N. Schwarz and F. Strack, grant SWF0044-6 from the Bundesminister für Forschung und Technologie of the Federal Republic of Germany to N. Schwarz, and ZUMA's "Cognition and Survey Research" program. We thank George Bishop for helpful comments on a previous draft.

Measurement Errors in Surveys.
Edited by Biemer, Groves, Lyberg, Mathiowetz, and Sudman.

ISBN: 0-471-53405-6

impact of the order in which response alternatives are presented and outline a cognitive model of response order effects.

3.2 RESPONSE ALTERNATIVES: WHAT THEY MAY TELL YOUR RESPONDENTS

The advantages and disadvantages of open- and closed-response formats have been the topic of considerable debate in survey research since its early days (e.g., Lazarsfeld, 1944; Krech and Crutchfield, 1948). As Converse (1984) noted, the debates were primarily based on "enlightened common sense, in-house experience, and practical constraints" (p. 279), rather than on systematic experimental research. As a result, the methodological debate lacked a coherent theoretical framework and the actual data base is still surprisingly small, given the importance of the issue.

Survey researchers typically assume that response alternatives constitute a "measurement device" that respondents use to report their answer. According to this assumption, respondents recall an opinion from memory, or compute it when asked, and select a response alternative to communicate their opinion to the researcher. As long as the response alternatives allow them to communicate their opinion, they are assumed to have little impact on the obtained results. Systematic bias is only expected if the response alternatives are too constrained, thus providing no opportunity to communicate the "true" answer. This assumption is captured in the concept of "question constraint" (see Schuman, 1985), which holds that respondents assume that they have to work within the set of response alternatives provided to them.

From a cognitive point of view, the concept of "question constraint" addresses only parts of the relevant processes. The survey interview is best considered as an ongoing conversation that includes the intertwined tasks of question comprehension, recall of information from memory, computation of an answer, and reporting of this answer to an interviewer (see Feldman, 1991; Strack and Martin, 1987; Strack and Schwarz, in press; Tourangeau and Rasinski, 1988, for more detailed discussions). Much as in other forms of social discourse in everyday life, all contributions of the participants to the conversation may influence each of these stages of the question answering process. In the survey interview, the contributions of the interviewer/researcher include the response alternatives provided to respondents, and respondents treat these contributions as they treat any other contribution to an ongoing conversation. That is, they proceed on the basis of the cooperativeness

principle that governs the conduct of conversation in everyday life (cf. Grice, 1975; Clark, 1985). This principle holds that every contribution should be relevant to the aim of the ongoing conversation, and that speakers should not provide information that is irrelevant to the task at hand. Moreover, speakers are required to make their contributions informative, that is, to provide information that the recipient needs, rather than information that the recipient already has — or may take for granted anyway. Conforming to these conversational norms requires a considerable degree of inference to determine which information is "informative" in the specific context given (cf. Schwarz, Strack, and Mai, 1991; Strack and Schwarz, in press). In the survey interview, this context is, in part, constituted by the response alternatives. Accordingly, response alternatives not only serve to record respondents' answers, but also to define respondents' substantive tasks. They therefore influence respondents' interpretation of the questions and determine which information they use in making a judgment and which responses they consider appropriate to report.

Whereas researchers in experimental (social) psychology pay close attention to the information that their subjects may extract from the research procedures used (cf. Wyer, 1974), the informative functions of apparently formal features of questionnaires have received little attention in social and psychological research. It is these informative functions that are of key interest in the first part of this chapter.

3.3 KNOWLEDGE AND OPINION QUESTIONS

3.3.1 Open- vs. Closed-Question Formats

Experimental studies on the impact of open- and closed-response formats in the domain of knowledge and opinion questions converge on the finding that open- and closed-response formats may yield considerable differences in the marginal distribution as well as in the ranking of items (e.g., Bishop, et al. 1988; Schuman, and Presser, 1977). On the one hand, any given opinion is less likely to be volunteered in an open-response format than to be endorsed in a closed-response format if presented. On the other hand, opinions that are omitted from the set of response alternatives in a closed format are unlikely to be reported at all, even if an "other" category is explicitly offered, which respondents in general rarely use (Bradburn, 1983; Molenaar, 1982). Several processes are likely to contribute to these findings.

First, precoded response alternatives may remind respondents of options that they may otherwise not consider. From a cognitive

perspective, open-response formats present a free-recall task to respondents, whereas closed formats present a recognition task. As has been found in other domains of research, recognition tasks have been found to result in higher degrees of recall (cf. Smyth, et al. 1987). For example, when respondents are asked, "What are the most important problems facing the country today?", a precoded list of response alternatives may direct respondents' attention to issues that may otherwise not have come to mind. When reminded of it, respondents may endorse the issue as important. Accordingly, Schuman and Presser (1977) observed, for example, that "security" was mentioned 13 percent less often as an important feature of a job when respondents were asked in an open rather than a closed format. Thus, open-response formats are more adequate when the investigator is interested in the salience of an issue, where the order in which a respondent retrieves different issues and the total number of respondents who retrieve a particular issue are of primary interest (cf. Bodenhausen and Wyer, 1987). Closed formats, on the other hand, are more appropriate when the investigator is interested in a fairly complete evaluation of a large set of issues to determine their relative importance.

Second, respondents are unlikely to report spontaneously, in an open-answer format, information that seems self-evident or irrelevant. In refraining from these responses they follow the conversational maxim that an utterance should be informative, as discussed above. This results in an underreporting of presumably self-evident information that is eliminated by closed-response formats, where the explicit presentation of the proper response alternative indicates the investigator's interest in this information.

In addition, respondents may frequently be uncertain if information that comes to mind does or does not belong to the domain of information the investigator is interested in. Again, closed-response formats may reduce this uncertainty, resulting in higher responses. This differential complexity of open and closed formats is reflected in higher nonresponse rates in the open-response form, in particular among less educated respondents (Schuman and Presser, 1981).

In combination, these processes are likely to result in an "underreporting" of opinions and knowledge in an open-response format as compared to a closed format. For behavioral reports, on the other hand, the pattern may reverse under some conditions. For example, Blair et al. (1977) found pronounced underreporting in response to threatening behavioral questions with a closed-answer format. This effect is most likely mediated by the informative function of response alternatives, as will be discussed below.

As a first conclusion, we note that open- and closed-response formats

pose different cognitive tasks and result in different patterns of responses, thus undermining the comparability of the data obtained under open- and closed-response format conditions. Next we consider specific types of response alternatives that are frequently used in survey research.

3.3.2 Middle Alternatives

Opinion questions often include a middle alternative between two extreme responses. This practice has been the topic of some debate. In general, the data indicate that explicitly offering the logically possible middle alternative will produce a considerable increment in the percentage of respondents who will endorse it (see Molenaar, 1982, for a review), as compared to conditions where the middle alternative has to be volunteered. Moreover, offering a middle alternative is likely to reduce the rate of "don't know" responses. Regarding the impact of middle alternatives on the endorsement of polar opposites, the findings are mixed. While Schuman and Presser (1981) found that the introduction of middle alternatives did not change the substantive conclusions drawn from the endorsement of the opposites, Bishop (1987) reported a series of studies where the introduction of middle alternatives affected the substantive conclusions. The conditions that determine whether middle alternatives draw similar or dissimilar numbers of respondents from both sides of the attitude continuum, thus affecting or not affecting the substantive conclusions, are a promising area for future research.

3.3.3 "Don't Know" Options and the Use of Filter Questions

Survey researchers frequently use "no opinion" filters to screen out respondents who may not have an opinion on the issue under investigation. This is either accomplished by offering a "no opinion" option as part of a set of precoded response alternatives (often referred to as a *quasi filter*), or by asking respondents if they have an opinion on the issue before the question proper is asked (often referred to as a *full filter*).

Research on the use of these filters (see Bishop et al. 1983; Schuman and Presser, 1981; Sudman and Bradburn, 1974, for reviews) indicates that respondents are more likely to endorse a "no opinion" option if it is explicitly offered than if they have to volunteer it. Moreover, they are more likely to report not having an opinion on the issue if a full filter is used than if a "no opinion" option is offered as part of the response

alternatives — and the more so, the more strongly the full filter is worded. For example, asking respondents if they have an opinion on the issue results in fewer "no opinion" responses than asking them if they "have thought enough about the issue to have an opinion on it" (Bishop et al. 1983).

Consistent with the informative functions approach offered here, a series of experiments demonstrated that filter questions influence respondents' perception of their task (Hippler and Schwarz, 1989): The more strongly the filter question is worded, the more respondents assume that they will have to answer difficult questions and that they may not have the required knowledge. Most importantly, full filters seem to suggest to respondents that they will *not* be asked a global opinion question, in contrast to actual survey practice. Accordingly, strongly worded filter questions discourage respondents from offering any global opinions that they may hold by suggesting a more demanding task than is actually in store. In line with this assumption, all respondents who reported not having an opinion in response to a filter question in one of our studies (Hippler and Schwarz, 1989, experiment 3) subsequently provided substantive responses on a global opinion question — presumably because the global question asked was less demanding than expected on the basis of the filter. As a result, the use of full filters seems likely to result in an underreporting of opinions.

3.3.4 Rating Scales

Like discrete response categories, rating scales are ubiquitous in social research, particularly in attitude measurement. Dawes and Smith (1985) provide a careful discussion of their properties and of the empirical and psychological justifications for their use. Leaving concerns about their psychometric properties aside (cf. Nunnally, 1978), rating scales with labeled endpoints do not seem to be very controversial. Respondents are able to use these scales consistently, even in telephone interviews without visual aids (cf. Hormuth and Brückner, 1985). Seven-point scales seem to be best in terms of reliability, percentage of undecided respondents, and respondents' ability to discriminate between the scale values (Cox, 1980). Thus, seven plus or minus two is the usual recommendation. Moreover, scales which provide verbal labels for each scale value seem more reliable than scales with labeled endpoints only (Krosnick and Berent, 1990).

Researchers should be aware, however, that the terms used to label the endpoints, and the terms used to designate the separate values of verbal rating scales, affect the obtained distribution (Rohrmann, 1978;

Wegner et al. 1982; Wildt and Mazis, 1978). Moreover, Schwarz, Knäuper et al. (in press) observed that respondents may use the specific numeric values provided by the researcher to interpret the meaning of the scale's labels. A representative sample of German adults were asked, "How successful would you say you have been in life?". The question was accompanied by an 11-point rating scale, ranging from "not at all successful" to "extremely successful." However, in one condition the numeric values of the rating scale ranged from 0 ("not at all successful") to 10 ("extremely successful"), whereas in the other condition they ranged from −5 ("not at all successful") to +5 ("extremely successful"). The results showed a dramatic impact of the numeric labels. Whereas 34 percent of the respondents endorsed a value between 0 and 5 on the 0 to 10 scale, only 13 percent endorsed one of the formally equivalent values between −5 and 0 on the −5 to +5 scale. Subsequent experiments indicated that this difference reflects differential interpretations of the term "not at all successful." When this label is combined with the numeric value 0, respondents interpret it to reflect the absence of success. However, when the same label is combined with the numeric value −5, they interpret it to reflect the presence of failure.

This differential interpretation of the same term as a function of its accompanying numeric value is also reflected in inferences that judges draw on the basis of a report given along a rating scale. For example, in one of our experiments, a fictitious student reported his academic success along one of the above scales, checking either a −4 or a 2. As expected, judges who were asked to estimate how often this student had failed an exam assumed that he failed twice as often when he checked a −4 than when he checked a 2, although both values are formally equivalent along 11-point rating scales of the type described above.

3.4 BEHAVIORAL QUESTIONS: WHAT RESPONDENTS LEARN FROM SCALES

So far, we have seen how the set of response alternatives provided by the researcher helps respondents to define their task. This, however, does not exhaust the informative functions of response alternatives. As a considerable body of research has demonstrated, response alternatives also inform respondents about the researcher's knowledge of, or assumptions about, the range of opinions or behaviors in the population, thus providing information about the "real world," which respondents may use in computing a judgment.

For example, research on the use of response alternatives in the assessment of behavioral frequency reports (see Schwarz, 1990a,b;

Table 3.1. Reported Daily TV Viewing as a Function of Response Alternatives

Reported Daily TV Viewing			
Low Frequency Alternatives		High Frequency Alternatives	
Up to 1/2 h	7.4%	Up to 2 1/2h	62.5%
1/2 h to 1h	17.7%	2 1/2h to 3h	23.4%
1h to 1 1/2h	26.5%	3h to 3 1/2h	7.8%
1 1/2h to 2h	14.7%	3 1/2h to 4h	4.7%
2h to 2 1/2h	17.7%	4h to 4 1/2h	1.6%
More than 2 1/2h	16.2%	More than 4 1/2h	0.0%

Note. $N = 132$. Adapted from Schwarz et al. (1985), "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments," *Public Opinion Quarterly*, 49, 388–395. Reprinted by permission.

Schwarz and Hippler, 1987, for reviews) indicates that respondents assume that the range of the response alternatives reflects the researcher's knowledge about the distribution of the behavior. Specifically, values in the middle range of the scale are assumed to reflect the "average" or "typical" behavior, whereas the extremes of the scale are assumed to correspond to the extremes of the distribution. These assumptions influence respondents' behavioral reports as well as related judgments in various ways.

3.4.1 Behavioral Reports

First, respondents use the range of the response alternatives as a frame of reference in estimating their own behavioral frequencies and report higher frequencies on scales that present high rather than low frequency response alternatives. The results of a study on TV viewing, shown in Table 3.1, illustrate this effect (Schwarz, et al. 1985). Whereas 37.5 percent of the respondents reported a daily viewing of 2 1/2 hours or more if presented the high frequency response alternatives, only 16.2 percent of the respondents did so if presented the low frequency response alternatives.

This reflects that individual instances of mundane and frequent behaviors, such as watching TV, are not separately represented in memory (see Bradburn et al. 1987; Schwarz, 1990a; and Strube, 1987, for reviews). Rather, individual episodes tend to blend into one generic representation, thus making it difficult to determine their frequency on the basis of a "recall the episodes and count their number" procedure (see Blair and Burton, 1987; Sudman and Schwarz, 1989). Accordingly, respondents have to rely on an estimation strategy, for which they will use any information that seems helpful (Bradburn et al., 1987). One piece

of information that is highly salient in the interview context is the range of the response alternatives provided. Respondents are therefore likely to use this range as a salient frame of reference in estimating their own behavioral frequency, resulting in higher estimates on high frequency compared to low frequency scales. The impact of this salient frame of reference is more pronounced when the episodic information that respondents can recall from memory is less. Accordingly, it is more pronounced for proxy reports than for self-reports (Schwarz and Bienias, 1990, experiments 1 and 2), and is eliminated when respondents have a chance to refresh their memory, e.g., by browsing through a TV program guide.

3.4.2 Comparative Judgments

The impact of the response alternatives is not limited to the behavioral frequency question with which they are provided. Given the assumption that the scale reflects the distribution of the behavior, checking a response alternative is the same as locating one's own position in the distribution. Accordingly, respondents extract comparison information from their own location on the response scale and use this information in making subsequent comparative judgments (e.g., Schwarz et al., 1985; Schwarz and Scheuring, 1988).

For example, respondents of the TV viewing study described above reported that TV plays a more important role in their leisure time (experiment 1), and described themselves as less satisfied with the variety of things they do in their leisure time (experiment 2), when they had to report their TV viewing on the low rather than on the high frequency scale. This reflects that checking a value in the upper range of the low frequency scale suggested to them that they watch *more* TV than "average," whereas their location on the high frequency scale suggested to them that they watch *less* TV than "average" (cf. Table 3.1).

Interestingly, the use of scale location in comparative judgments is not limited to respondents, but does affect the users of their reports as well. For example, experienced physicians were more likely to consider that having a given physical symptom "twice a week" reflected a serious medical condition if that frequency were reported on a low rather than a high frequency scale (Schwarz, Bless et al., 1991).

3.4.3 Question Interpretation

Finally, if the target behavior is open to interpretation, as is often the case when subjective experiences are assessed, respondents may use the frequency range of the response alternatives to determine the exact

reference of the question. For example, respondents who were asked to report how frequently they feel "really annoyed" on a low frequency scale subsequently reported more extreme annoyances as typical for their experience than respondents who had to give their frequency report on a high frequency scale (Schwarz et al. 1988). This suggests that respondents used the frequency range of the response alternatives to determine what the researcher meant by "really annoyed," asking themselves "Does that refer to major or to minor annoyances?" Given that minor annoyances are frequent, whereas major ones are rare, the frequency range of the response alternatives helped to clarify the meaning of the question. Accordingly, the same behavioral question in combination with different frequency alternatives is likely to assess different experiences.

3.5 SUMMARY: THE INFORMATION FUNCTION OF QUESTIONNAIRES

We conclude from these and related findings that respondents actively use apparently "formal" features of the questionnaire as a source of information to determine what is expected of them and to solve the cognitive tasks posed in the survey interview (see Schwarz, 1990a; Schwarz and Strack, 1991; Strack and Martin, 1987; and Strack and Schwarz, in press, for more extended discussions). Accordingly, response alternatives are not only technical measurement devices. They are better conceptualized as part of an ongoing conversation between the researcher and the respondent. Although survey respondents are aware that they are supposed to answer rather than to ask questions, they still bring many assumptions to the survey interview that govern the conduct of conversations in everyday life (cf. Grice, 1975). Most importantly, they expect all participants to provide only information that is relevant to the issue at hand, and they assume that the selection and presentation of response alternatives comply with that norm. As a result, they conduct their own share of thinking, judging and communicating within the framework of the ongoing conversation, much as they would be expected to do in everyday life — except that in the survey interview that context is prestructured by the questionnaire. For that very reason, we have to pay close attention to the context that we set up in devising the questionnaire, unless we want to assess what we evoked in the first place.

3.6 THE IMPACT OF RESPONSE ORDER: ACCOUNTING FOR PRIMACY AND RECENCY

Once the researcher decides which response alternatives should be selected, he or she needs to determine the order in which they are

presented to respondents. Again, the survey literature indicates that the presentation order may strongly influence the obtained results. Theoretically, *primacy effects*, that is, higher endorsements of items presented early in the list, as well as *recency effects*, that is, higher endorsements of items presented late in the list, may be obtained. While response order effects have occasionally been reported when the response alternatives present an ordered set of categories that constitute a verbal rating scale (e.g., excellent, very good, good, fair, poor), they are rare under these conditions (see Mingay and Greenwell, 1989). In contrast, response order effects have frequently been obtained when each response alternative presents a different opinion on an issue, and respondents are asked to select the one that best represents their own position. Several processes are likely to contribute to these findings, as discussed below. Moreover, these processes may result in complex interaction effects, depending upon the specific conditions. Accordingly, this area of research is characterized by a complex set of apparently contradictory findings.

3.6.1 Elaboration Processes

As one heuristic framework for understanding the nature of response order effects, we suggest that each response alternative may be portrayed as a single persuasive argument (Schwarz, Hippler et al. 1991; see also Krosnick, 1991a, and Krosnick and Alwin, 1987, for a related conceptualization, emphasizing memory limitations). Borrowing from research on the processing of persuasive communications (see Petty and Cacioppo, 1986 for a detailed review), we assume that a given item is more likely to be endorsed the more positive cognitive responses it elicits, that is, the more agreeing thoughts the respondent generates. Conversely, a given item should be less likely to be endorsed the more disagreeing thoughts it elicits. The number of thoughts generated, however, is not only a function of the content of the item per se, but also a function of the degree of cognitive elaboration that a given mode of data collection permits.

Suppose, for example, that a long list of response alternatives is presented to respondents on a show card as part of a face to face interview, or in a self-administered questionnaire. Under these conditions, "items presented early in a list are likely to be subjected to deeper cognitive processing," as Krosnick and Alwin (1987, p. 213) noted. "By the time a respondent considers the later alternatives, his or her mind is likely to be cluttered with thoughts about previous alternatives that inhibit extensive consideration of later ones." Accordingly, a given response alternative is more likely to be endorsed if presented early rather than late in the list, provided that it is "plausible" to the respondent, thus eliciting agreeing thoughts. Conversely, an "implau-

sible" response alternative that elicits disagreeing thoughts is less likely to be endorsed if presented early.

Note that this analysis implies that order effects are likely to depend on respondents' attitudes, which determine what is "plausible" for a given respondent. For example, a "liberal" item is likely to elicit agreeing thoughts from liberal respondents, but disagreeing thoughts from conservative respondents. Accordingly, liberal respondents should be *more* likely to endorse the item if presented early rather than late on a list, whereas conservative respondents should be *less* likely to endorse it under these conditions. As a result, primacy and recency effects in subsamples may cancel one another, resulting in the apparent absence of order effects in the sample as a whole. Hippler, et al. (1990) report data supporting this hypothesis.

Assume, however, that the items are not presented visually, but are read to respondents by the interviewer. In this case, respondents have little opportunity to elaborate on the items presented early in the list, because the time that is available for processing each item is restricted by the speed with which the interviewer moves on to read the next one. "Under these circumstances, respondents are able to devote most processing time to the *final* item(s) read, since interviewers usually pause most after reading them" (Krosnick and Alwin, 1987, p. 203). In addition, respondents may find it difficult to keep all response alternatives in mind without visual help. Accordingly, items that elicit agreeing thoughts from a given respondent should be more likely to be endorsed if presented late rather than early in the list, resulting in recency effects under auditory presentation formats. Again, the reverse holds for items that elicit disagreeing thoughts.

In summary, response order effects are assumed to depend on the items' serial position, their plausibility for a given respondent, and the administration mode used. If the response alternatives are presented on show cards or in a self-administered questionnaire, items presented early in the list are more likely to be extensively processed than items presented later, resulting in *primacy effects*, provided that the item is plausible to the respondent (i.e., that it elicits agreeing thoughts). In contrast, if the items are read to respondents, the last response alternatives aremore likely to be extensively processed and recalled than the first ones, resulting in *recency effects*, again assuming plausibility of the items for the respondent. Given that the likelihood of endorsement may be expected to decrease as more extensive processing uncovers flaws in implausible items, the reverse predictions hold for items that lack plausibility for the respondent (i.e., items that elicit disagreeing thoughts; see Schwarz, Strack, and Mai, 1991, for further discussion).

The predicted interaction of serial position and administration mode

for plausible items has been supported by secondary analyses of a large number of split-ballot experiments with representative samples of the adult population in West Germany, which were originally conducted by the Allensbach Institute under the direction of Elizabeth Noelle-Neumann, since the early 1950s (see Schwarz, Hippler, and Noelle-Neumann, 1991, for a review). Additional analyses of these archival data, using respondents' attitudes as an approximation for the likely plausibility that a given item may have for them, provided preliminary support for the assumed role of item plausibility (Hippler et al. 1990). We emphasize, however, that our conclusions are based on secondary analyses involving different questions under different presentation formats. While the consistency of the data patterns across widely different questions suggests that the conclusions are likely to be valid, more tightly controlled experiments using the same questions and comparable samples under all conditions are definitely needed.

3.6.2 Memory Limitations

Not surprisingly, response order effects have frequently been attributed to respondents' memory limitations, and Krosnick and Alwin (1987; see also Krosnick, 1991a) suggested that the interaction of serial position and administration mode described above may reflect memory processes. Despite their popularity in the survey literature, however, memory based accounts are difficult to reconcile with the available data.

As psychological research on the learning of long lists of verbal expressions indicates, the recall of verbal material depends on its serial position in the list and the time delay between learning and testing (see Smyth, et al., 1987, for a review). Material that is presented at the beginning of the list is more likely to enter long-term memory than material that is presented later, because the first few items "suffer less competition for time and space in immediate memory from other items" (Smyth, et al., 1987, p. 123). On the other hand, material that is presented at the end of the list may still be in short-term memory if recall follows learning without much delay. This results in an interaction effect of serial position and the delay between learning and recall: Without delay, material presented at the end of the list is much more likely to be recalled than material presented at the beginning, reflecting that the later items can be recalled from short-term memory. Accordingly, recency effects are typically obtained under no-delay conditions, and they are more pronounced if the material is read to subjects rather than presented visually (Murdock and Walker, 1969). If recall is delayed, however, material presented at the beginning of the list is more likely to be

remembered than material presented at the end, reflecting that the early items entered long-term memory, whereas the later items can no longer be recalled from short-term memory. This results in primacy effects under delayed recall conditions. Material that is presented in the middle of the list is least likely to be recalled under any conditions.

How do these findings bear on response order effects in survey measurement? Given that respondents report their answers immediately after exposure to the response alternatives, no delay between "learning" and "recall" is introduced. Accordingly, response alternatives presented at the end of the list should be easily accessible in short-term memory. Because recall of late items from short-term memory is better than recall of early items from long-term memory under no-delay conditions, this should result in pronounced recency effects. Primacy effects should only be obtained if a delay is introduced between exposure to the response alternatives and respondents' reports. This is typically *not* the case in survey interviews. Nevertheless, *primacy* rather than recency effects have typically been reported in survey experiments with long lists of response alternatives (e.g., Payne, 1951; Mueller, 1970; Ring, 1975; Krosnick and Alwin, 1987), suggesting that memory limitations are not the primary source of response order effects in survey measurement.

In fact, long lists of response alternatives are usually presented on show cards that remain available until respondents report their answer, thus placing little burden on their memory to begin with. Moreover, response order effects have consistently been observed on questions that present only two or three response alternatives (e.g., Payne, 1951; Schuman and Presser, 1981; Schwarz, Hippler, et al. in press). This limited number of response alternatives, however, should be easily accessible in short-term memory under the no-delay condition of survey interviews, thus rendering the emergence of memory-based order effects unlikely unless the alternatives are overly complex. We conclude from these inconsistencies that memory limitations are *not* the dominant source of response order effects in survey measurement; whether the cognitive response approach that we offered above, building on Krosnick and Alwin's (1987) discussion, fares much better, on the other hand, remains to be seen.

3.6.3 Contrast Effects

Finally, the likelihood that a given response alternative is endorsed does not only depend on its serial position per se, which may influence its cognitive elaboration or memorability, but also on the nature of the preceding response alternatives. This possibility has received little

attention in survey research. Specifically, if a given item is preceded by an item that is more extreme on the dimension of judgment, a *contrast effect* may emerge, provided that all items are to be judged along the same dimension (cf. Ostrom and Upshaw, 1968).

Suppose, for example, that respondents are asked to select persons that they like well, and that an extremely well-liked person is presented in the middle of a list. If so, a moderately liked person will seem *less* likable if he or she is presented in the second half of the list, following the extreme stimulus, than if he or she is presented at the beginning of the list, preceding the extreme stimulus. If we compared the two orders of this list, this judgmental contrast effect would lead us to conclude that a pronounced primacy effect emerged. On the other hand, if the person presented in the middle of the list were extremely dislikable, the same mechanism of judgmental contrast would increase the endorsement of moderately liked persons presented in the second half of the list. In that case, a comparison of both order conditions would lead us to conclude that a pronounced recency effect emerged. Note, however, that the underlying cognitive process of judgmental contrast is quite different from the cognitive elaboration and memory processes discussed above.

A classic example of such a contrast effect was reported by Noelle-Neumann (1970). Specifically, respondents were presented a list of food items and were asked to select the ones that are typically "German." Respondents were more likely to consider a number of food items, such as noodles or potatoes, as typically "German" when they were preceded by rice than when they were not. Thus, introducing rice as the first item resulted in pronounced contrast effects in the perception of the other food items. Finally, the evaluation of rice itself was unaffected by order manipulations.

Contrast effects of this type are a function of the items' extremity on the underlying dimension of judgment. Introducing a more extreme item results in a wider "perspective" regarding the set of stimuli, thus affecting the evaluation of moderate stimuli as described in Ostrom and Upshaw's (1968) perspective theory. Accordingly, these effects do also emerge under conditions where each item is likely to receive about the same degree of attention and elaboration, for example, because *each* item has to be rated along some scale. For example, Schwarz et al. (1990) observed that subjects evaluated a number of drinks as being more "typically German" when they were preceded by "vodka" (an extremely atypical drink) than when they were preceded by "beer" (an extremely typical drink).

Moreover, contrast effects of this type do *not* require that the items are presented on the same list. Rather, they may also emerge if the extreme item is presented as part of a preceding question, provided that

this question taps the same dimension of judgment. For example, in some conditions of the Schwarz, et al. (1990) study, we asked some respondents to estimate the percentage of Germans who drink vodka, and others to estimate the percentage of Germans who drink beer, before they rated the typicality of other drinks. As expected, subjects who estimated the percentage of Germans who drink vodka rated subsequent drinks as more typically German than subjects who estimated how many Germans drink beer, replicating the contrast effects obtained when all stimuli were presented on the same list. Other subjects, however, were asked as part of the preceding questions to estimate the caloric content, rather than the consumption, of vodka or beer. While this question also serves to render these drinks highly salient in the interview context, it does not tap the typicality dimension that underlies estimates of the consumption of these drinks. Accordingly, estimating their caloric content did *not* influence subsequent typicality ratings. Thus, we conclude that contrast effects can emerge as a function of preceding questions *if* these questions tap the same underlying dimension of judgment.

This emergence of contrast effects bears in important ways on the emergence of primacy and recency effects in general: If an extremely *positive* item is presented as part of the stimulus set, it will *decrease* the endorsement of subsequent moderate items. If an extremely *negative* item is presented, on the other hand, it will *increase* the endorsement of subsequent moderate items. These judgmental effects may lead the researcher to conclude that the data show pronounced recency or primacy effects. Accordingly, the phenomenon of judgmental contrast may dilute the emergence of memory and elaboration phenomena.

3.6.4 Summary of Response Order Effects

As the preceding discussion illustrates, response order effects may be a function of the serial position of the item, the item's plausibility for a given respondent, the extremity of adjoining items, and the administration mode used. The effects become complex when several of these factors are simultaneously present, as discussed in detail by Schwarz, Hippler, and Noelle-Neumann (1991). The absence of significant response order effects may indicate that none of the factors causing them was sufficiently powerful in a given case, but it may also reflect the different effects canceling each other. Because the relative power of the factors involved is difficult to evaluate a priori, the emergence of response order effects is likely to remain a surprise in many specific cases, although the underlying processes seem systematic and their effects replicable.

ZUMA-Arbeitsberichte

80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA-Arbeitstagung vom 25.09. - 05.10.79

81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter Schmidt
Methodenbericht zum "ALLBUS 1980"

81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman
Integration von qualitativen und quantitativen Forschungsansätzen

82/03 Gerhard Arminger, Horst Busse, Manfred Küchler
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung

82/08 Glenn R. Carroll
Dynamic analysis of discrete dependent variables: A didactic essay

82/09 Manfred Küchler
Zur Messung der Stabilität von Wählerpotentialen

82/10 Manfred Küchler
Zur Konstanz der Recallfrage

82/12 Rolf Porst
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu einer Kritik des Fragenprogramms

82/13 Peter Ph. Mohler
SAR - Simple AND Retrieval mit dem Siemens-EDT-Textmanipulationsprogramm

82/14 Cornelia Krauth
Vergleichsstudien zum "ALLBUS 1980"

82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer
Methodenbericht zum "ALLBUS 1982"

83/09 Bernd Wegener
Two approaches to the analysis of judgments of prestige: Interindividual differences and the general scale

83/11 Rolf Porst
Synopse der ALLBUS-Variablen. Die Systematik des ALLBUS-Fragenprogramms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und ALLBUS 1982

84/01 Manfred Küchler, Peter Ph. Mohler
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten, Teil I + II

84/02 Bernd Wegener
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-Prestige-Skala

84/03 Peter Prüfer, Margrit Rexroth
Erfahrungen mit einer Technik zur Bewertung von Interviewerverhalten

84/04 Frank Faulbaum
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) 1982

84/05 Jürgen Hoffmeyer-Zlotnik
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten

84/07 Gabriele Hippler, Hans-Jürgen Hippler
Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the-Face" Technique

85/01 Hartmut Esser
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von Antwortverzerrungen in Interviews

85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang
Methodenbericht zum "ALLBUS 1984"

86/01 Dagmar Krebs
Zur Konstruktion von Einstellungsskalen im interkulturellen Vergleich

86/02 Hartmut Esser
Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung

86/03 Bernd Wegener
Prestige and Status as Function of Unit Size

86/04 Frank Faulbaum
Very Soft Modeling: The Logical Specification and Analysis of Complex Process Explanations with Arbitrary Degrees of Underidentification and Variables of Arbitrary Aggregation and Measurement Levels

86/05 Peter Prüfer, Margrit Rexroth (Übersetzung: Dorothy Duncan)
On the Use of the Interaction Coding Technique

86/06 Hans-Peter Kirschner
Zur Kessler-Greenberg-Zerlegung der Varianz der Meßdifferenz zwischen zwei Meßzeitpunkten einer Panel-Befragung

86/07 Georg Erdmann
Ansätze zur Abbildung sozialer Systeme mittels nicht-linearer dynamischer Modelle

86/09 Heiner Ritter
Einige Ergebnisse von Vergleichstests zwischen den PC- und Mainframe-Versionen von SPSS und SAS

86/11 Günter Rothe
Bootstrap in generalisierten linearen Modellen

87/01 Klaus Zeifang
Die Test-Retest-Studie zum ALLBUS 1984 - Tabellenband

87/02 Klaus Zeifang
Die Test-Retest-Studie zum ALLBUS 1984 - Abschlußbericht

87/04 Barbara Erbslöh, Michael Wiedenbeck
Methodenbericht zum "ALLBUS 1986"

87/05 Norbert Schwarz, Julia Bienias
What Mediates the Impact of Response Alternatives on Behavioral Reports?

87/06 Norbert Schwarz, Fritz Strack, Gesine Müller, Brigitte Chassein
The Range of Response Alternatives May Determine the Meaning of the Question: Further Evidence on Informative Functions of Response Alternatives

87/07 Fritz Strack, Leonard L. Martin, Norbert Schwarz
The Context Paradox in Attitude Surveys: Assimilation or Contrast?

87/08 Gudmund R. Iversen
Introduction to Contextual Analysis

87/09 Seymour Sudman, Norbert Schwarz
Contributions of Cognitive Psychology to Data Collection in Marketing Research

87/10 Norbert Schwarz, Fritz Strack, Denis Hilton, Gabi Naderer
Base-Rates, Representativeness, and the Logic of Conversation

87/11 George F. Bishop, Hans-Jürgen Hippler, Norbert Schwarz, Fritz Strack
A Comparison of Response Effects in Self-Administered and Telephone Surveys

87/12 Norbert Schwarz
Stimmung als Information. Zum Einfluß von Stimmungen und Emotionen auf evaluative Urteile

88/01 Antje Nebel, Fritz Strack, Norbert Schwarz
Tests als Treatment: Wie die psychologische Messung ihren Gegenstand verändert

88/02 Gerd Bohner, Herbert Bless, Norbert Schwarz, Fritz Strack
What Triggers Causal Attributions? The Impact of Valence and Subjective Probability

88/03 Norbert Schwarz, Fritz Strack
The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction

88/04 Hans-Jürgen Hippler, Norbert Schwarz
"No Opinion"-Filters: A Cognitive Perspective

88/05 Norbert Schwarz, Fritz Strack
Evaluating One's Life: A Judgment of Subjective Well-Being

88/06 Norbert Schwarz, Herbert Bless, Gerd Bohner, Uwe Harlacher, Margit Kellenbenz
Response Scales as Frames of Reference:
The Impact of Frequency Range on Diagnostic Judgments

88/07 Michael Braun
Allbus-Bibliographie (7. Fassung, Stand: 30.6.88)

88/08 Günter Rothe
Ein Ansatz zur Konstruktion inferenzstatistisch verwertbarer Indices

88/09 Ute Hauck, Reiner Trometer
Methodenbericht
International Social Survey Program - ISSP 1987

88/10 Norbert Schwarz
Assessing frequency reports of mundane behaviors:
Contributions of cognitive psychology to questionnaire construction

88/11 Norbert Schwarz, B. Scheuring (sub.)
Judgments of relationship satisfaction: Inter- and intraindividual comparison strategies as a function of questionnaire strcuture

88/12 Rolf Porst, Michael Schneid
Ausfälle und Verweigerungen bei Panelbefragungen
- Ein Beispiel -

88/13 Cornelia Züll
SPSS-X. Anmerkungen zur Siemens BS2000 Version

88/14 Michael Schneid
Datenerhebung am PC - Vergleich der Interviewprogramme "interv^{+}" und "THIS"

88/15 Norbert Schwarz, Bettina Scheuring
Die Vergleichsrichtung bestimmt das Ergebnis von Vergleichsprozessen:
Ist - Idealdiskrepanzen in der Partnerwahrnehmung

89/01 Norbert Schwarz, George F. Bishop, Hans-J. Hippler, Fritz Strack
Psychological Sources Of Response Effects in Self-Administered And Telephone Surveys

89/02 Michael Braun, Reiner Trometer, Michael Wiedenbeck,
Methodenbericht. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften - ALLBUS 1988 -

89/03 Norbert Schwarz
Feelings as Information:
Informational and Motivational Functions of Affective States

89/04 Günter Rothe
Jackknife and Bootstrap:
Resampling-Verfahren zur Genauigkeitsschätzung von Parameterschätzungen

89/05 Herbert Bless, Gerd Bohner, Norbert Schwarz und Fritz Strack
Happy and Mindless?
Moods and the Processing of Persuasive Communications

89/06 Gerd Bohner, Norbert Schwarz und Stefan E. Hormuth
Die Stimmungs-Skala: Eine deutsche Version des "Mood Survey" von Underwood und Froming

89/07 Ulrich Mueller
Evolutionary Fundamentals of Social Inequality, Dominance and Cooperation

89/08 Robert Huckfeldt
Noncompliance and the Limits of Coercion:
The Problematic Enforcement of Unpopular Laws

89/09 Peter Ph. Mohler, Katja Frehsen und Ute Hauck
CUI - Computerunterstützte Inhaltsanalyse
Grundzüge und Auswahlbibliographie zu neueren Anwendungen

89/10 Cornelia Züll, Peter Ph. Mohler
Der General Inquirer III -
Ein Dinosaurier für die historische Forschung

89/11 Fritz Strack, Norbert Schwarz, Brigitte Chassein, Dieter Kern, Dirk Wagner
The Salience of Comparison Standards and the Activation of Social Norms: Consequences for Judgments of Happiness and their Communication

89/12 Jutta Kreiselmaier, Rolf Porst
Methodische Probleme bei der Durchführung telefonischer Befragungen: Stichprobenziehung und Ermittlung von Zielpersonen, Ausschöpfung und Nonresponse, Qualität der Daten.

89/13 Rainer Mathes
Modulsystem und Netzwerktechnik.
Neuere inhaltsanalytische Verfahren zur Analyse von Kommunikationsinhalten.

89/14 Jutta Kreiselmaier, Peter Prüfer, Margrit Rexroth
Der Interviewer im Pretest.
Evaluation der Interviewerleistung und Entwurf eines neuen Pretestkonzepts. April 1989.

89/15 Henrik Tham
Crime as a Social Indicator.

89/16 Ulrich Mueller
Expanding the Theoretical and Methodological Framework of Social Dilemma Research

89/17 Hans-J. Hippler, Norbert Schwarz, Elisabeth Noelle-Neumann
Response Order Effects in Dichotomous Questions:
The Impact of Administration Mode

89/18 Norbert Schwarz, Hans-J. Hippler, Elisabeth Noelle-Neumann, Thomas Münkel
Response Order Effects in Long Lists:
Primacy, Recency, and Asymmetric Contrast Effects

89/19 Wolfgang Meyer
Umweltberichterstattung in der Bundesrepublik Deutschland

89/20 Michael Braun, Reiner Trometer
ALLBUS Bibliographie (8. Fassung, Stand: 30.6. 1989)

89/21 Günter Rothe
Gewichtungen zur Anpassung an Statusvariablen.
Eine Untersuchung am ALLBUS 1986

89/22 Norbert Schwarz, Thomas Münkel, Hans-J. Hippler
What determines a "Perspective"?
Contrast Effects as a Function of the Dimension
Tapped by Preceding Questions

89/23 Norbert Schwarz, Andreas Bayer
Variationen der Fragenreihenfolge als Instrument
der Kausalitätsprüfung: Eine Untersuchung zur Neutralisationstheorie devianten Verhaltens

90/01 Norbert Schwarz, Fritz Strack, Hans-Peter Mai
Assimilation and Contrast Effects in Part-Whole
Question Sequences:
A Conversational Logic Analysis

90/02 Norbert Schwarz, Fritz Strack, Hans-J. Hippler, George Bishop
The Impact of Administration Mode on Response Effects in
Survey Measurement

90/03 Norbert Schwarz, Herbert Bless, Gerd Bohner
Mood and Persuasion: Affective States Influence the
Processing of Persuasive Communications

90/04 Michael Braun, Reiner Trometer
ALLBUS-Bibliographie 90

90/05 Norbert Schwarz, Fritz Strack
Context Effects in Attitude Surveys:
Applying Cognitive Theory to Social Research

90/06 Norbert Schwarz, Herbert Bless, Fritz Strack,
Gisela Klumpp, Annette Simons
Ease of Retrieval as Information:
Another Look at the Availability Heuristic

90/07 Norbert Schwarz, Fritz Strack, Hans-J. Hippler
Kognitionspsychologie und Umfrageforschung:
Themen und Befunde eines interdisziplinären Forschungsgebietes